CHAPTER II

LITERATURE REVIEWS

2.1 Structure and function of hemoglobin

Hemoglobin is the principal protein of the red blood cell. It is the most abundant in human and represents more than 95% of the soluble protein content of the erythrocytes. The major role of the hemoglobin is the transport of oxygen from the lungs to the body tissues; an associated function is the binding of carbon dioxide and protons by deoxyhemoglobin, thereby serving to buffer the blood on the venous side of the circulation (Honig and Adam III, 1986). Hemoglobin synthesis requires the coordinated production of heme and globin. The normal hemoglobin molecule contains two distinct globin chains. One of the chains is designated alpha and contains 141 amino acid residues. The second chain is called beta and contains 146 amino acid residues. In the normal adult, there are two α-polypeptide chains and two β-polypeptide chains, designated $\alpha_2\beta_2$ or hemoglobin A. The β-globin chain contains 8 helical segments, designated A-H (Figure 4).

2.2 Types of hemoglobin

All of the molecular forms of human hemoglobin (Table 1) share similar chemical and functional properties, but each also has a number of unique characteristics. These different kinds of hemoglobin, which are characteristic of different stages of development (embryonic, fetal and postnatal), differ in the composition of their chains: the hemoglobin of adults (Hb A) contains two α- and β-chains ($\alpha_2\beta_2$) normally makes up 94-97% of the hemoglobin in the circulating erythrocytes of adults. A small proportion (2-3%) of adult hemoglobin has two α- and δ-chains (HbA_2:$\alpha_2\delta_2$). Hemoglobin formed during the fetal period (Hb F) contains two α- and γ- chains ($\alpha_2\gamma_2$). In the embryonic stage,ζ-chain are joined to ε- or γ-chains (Hb Gower 1: $\zeta_2\varepsilon_2$ and Hb Portland: $\zeta_2\gamma_2$); two ε chain form Hb Gower 2 ($\alpha_2\varepsilon_2$).

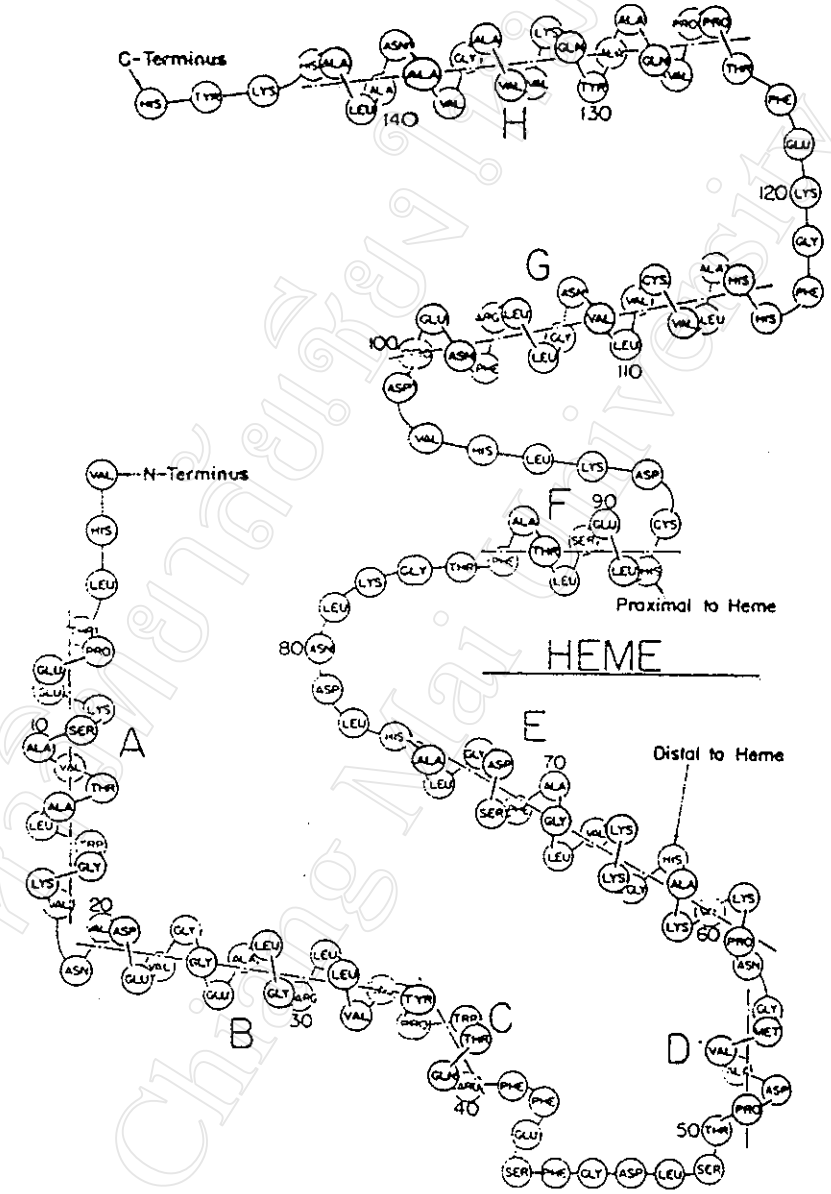

Figure 4. Representation of the secondary structure of the human hemoglobin β-globin chain, showing the helical and non-helical regions of the globin chain (Honig and Adams III, 1986).

Table 1. The normal human hemoglobins and their globin subunits (Honig and Adams III, 1986).

Hemoglobin	α-like subunit	β-like subunit	Tetramer composition
Hb A	α	β	$\alpha_2\beta_2$
Hb A_2	α	δ	$\alpha_2\delta_2$
Hb F	α	Gγ, Aγ	$\alpha_2\gamma_2$
Hb Gower-2	α	ε	$\alpha_2\varepsilon_2$
Hb Gower-1	ζ	ε	$\zeta_2\varepsilon_2$
Hb Portland	ζ	γ	$\zeta_2\gamma_2$

2.3 The localization and organization of globin gene

The alpha and beta globin genes were inherited independent and there by occupied the sites on separate chromosome (Smith and Torbert, 1958). The human hemoglobin genes exist as two separate clusters of related multigene families, a frequent type of organization of mammalian genes. The alpha gene cluster is located on the short arm of chromosome 16 over a 25-kb region. The beta gene cluster is situated on the short arm of another chromosome, 11, across a 60-kb region (Diesseroth *et al.*, 1976,1977,1978).

2.3.1 The α-globin gene cluster

The structural genes of the α-gene cluster include two α-globin genes (α_1 and α_2). A ζ-gene is active only during the embryonic period. Three pseudogene: $\psi\zeta$, $\psi\alpha_2$ and $\psi\alpha_1$ are located in between. A further gene, θ, with unknown function, has been identified in this region. These gene maps shows the globin gene to be arranged in the order 5'- ζ - $\psi\zeta_1$ - $\psi\alpha_2$ - $\psi\alpha_2$- α_2 - α_1 - θ - 3' (Lauer *et al.*,1980).

2.3.2 The β-globin gene cluster

The β-globin gene cluster is located on the short arm of chromosome 11 in region 1, band 5.5 (11p15.5) (Figure 5). They span about 60,000 bp of DNA. There are two γ genes, $^{A}\gamma$ and $^{G}\gamma$, which differ only in codon 136. Codon 136 of $^{A}\gamma$ is alanine, and of $^{G}\gamma$ is glycine. A beta pseudogene ($\psi\beta_1$) is located between the $^{A}\gamma$ gene and the δ-gene. It is similar to the β-gene, but has been permanently altered by deletion and an internal stop codon, so that it cannot code for a functional polypeptide. A region that jointly regulates these genes is located upstream (in the 5' direction) from the β-gene. The gene map of the β-globin gene cluster that was developed by Fritsch *et al.*(1980) also contains, in addition to the ε -γ -, δ -, and β-globin genes, a β-globin-like sequence. The β-globin-like sequence was identified at a position approximately intermediate between the $^{A}\gamma$ and δ-genes (Figure 6) may be a "pseudogene" a vestigial structure that presumably arose by gene duplication and which may have been expressed as a functional globin gene in an earlier evolutionary period. An additional β-like pseudogene was also believed to exist in a position 5' to the ε-gene. But a subsequent detailed structural analysis of this region showed no convincing evidence of globin like sequences (Shen and Smithies, 1982). The structural sequence of β-globin genes included three exons separated by two introns (Lawn *et al.*,1978,1980). A "promoter" region located 5' to the coding portion of the gene contains two and possibly more of these sequences (Figure 7). This region is beliveved to be the recognition and binding site for RNA polymerase, the enzyme which catalyzes mRNA systhesis. The first of these promoter sequence, the ATA box or Goldberg Hogness box (Proudfoot, 1979) is located 29-30 bp 5' to the transcribed portion of the genes. The function of this sequence appears to be to locate and direct the precise site of initiation of transcription of the gene (Mathis and Chambon, 1981). A second conserved sequence, the CCAAT box, is located 70-80 bp 5' to the transcription initiation site. A highly conserved hexanucleotide sequence AATAAA has also been identified in the 3' noncoding region of the globin genes, 15-19 bp 5' from the point of transcription termination (Michelson and Orkin, 1980; Efstratiadis *et al.*, 1980). The function of this sequence as a termination signal for

RNA polymerase, or as a recognition site for enzymes involved in the polyadenylation of mRNA (Proudfoot and Brownlee, 1976).

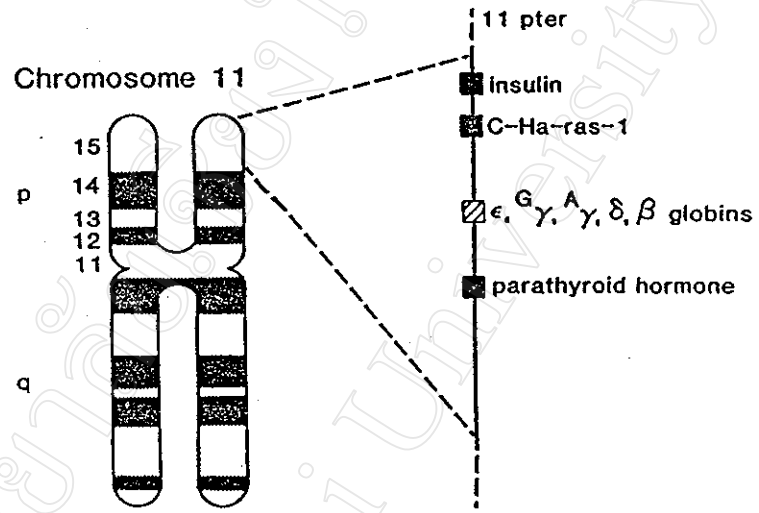

Figure 5. The location of the β-globin gene cluster on chromosome 11 (Honig and Adams III, 1986).

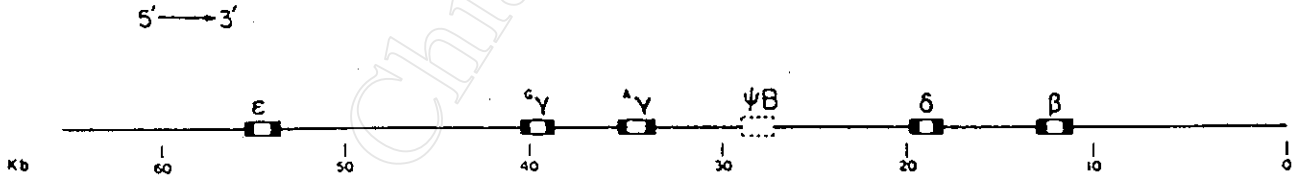

Figure 6. Linkage map of the human β-globin gene cluster. The location of the ε-, $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes and a β pseudogene (ψβ) are shown (Honig and Adams III, 1986).

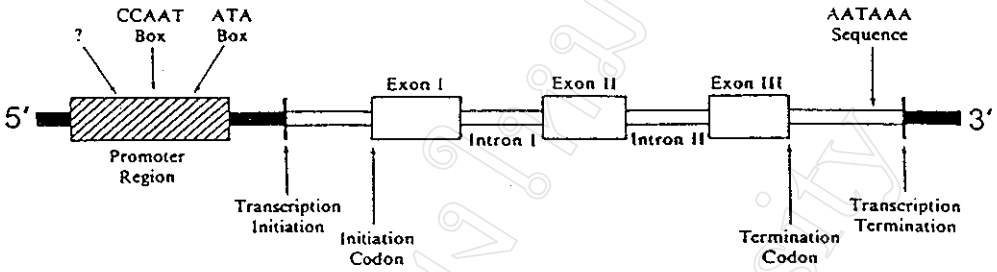

Figure 7. The general structure and organization of the globin genes. The transcribed portion of the gene is represented by the unshaded rectangular segments (Honig and Adams III, 1986).

2.4 The β-thalassemia

The thalassemia is a group of disorders in which the normal hemoglobin protein is absent or produced in reduced amounts. It affects either the β-chain (β-thalassemia) or the α-chain (α-thalassemia). Thalassemia is a complex group of disorders because of the genetics of hemoglobin production and the structure of the hemoglobin molecule. Thalassemia and hemoglobinopathies are widespread, recessive inherited diseases; approximately 250 million people (4.5% of the world population) carry abnormal hemoglobin genes. If one of the β-globin genes fails (for instance,b1 in Figure 8), the amount of β-globin in the cell is reduced by half. This situation is called thalassemia trait or the *thalassemia minor*. If both genes fail (b1 and b2 in Figure 8), no β-gobin chain is produced. This is called *thalassemia major*. In some cases, the genes failure is not total. The gene produces a small amount of normal β-globin chain. Sometimes, a person inherits two β-thalassemia genes in which the production of β-globin chain from each is reduced, but is not zero. The resulting clinical condition is more severe than thalassemia minor, where one gene fails but the other works normally. The condition is less severe than thalassemia major, where both β-globin genes fail completely. The clinical condition is termed, *thalassemia intermediate*.

In Thailand β-thalassemia is found at high frequencies in the North (5-8%). The molecular abnormalities that produce β-thalassemia comprise an unexpectedly large and diverse group of mutations. These various mutations, however, have a quite limited range of expression, with many of the different molecular abnormalities producing syndromes with indistinguishable clinical and hematological features. Virtually all forms of β-thalassemia are accompanied by the presence of small, under hemoglobinized erythrocytes, and at least some degree of anemia. Compensatory increases in the levels of Hb F and Hb A_2 are findings that are also common to most of these syndromes. In general the β^+-thalassemia have a milder clinical course as compared with the β°-thalassemia, but this distinction is often not a clear one, and some β^+-thalassemia mutations are known to produce quite severe clinical manifestations.

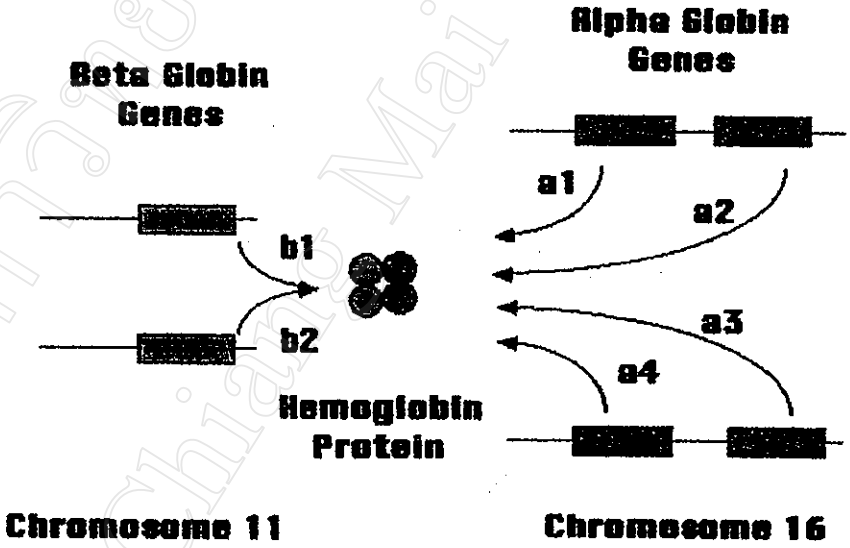

Figure 8. The two chromosomes 11 have one beta globin gene each (for a total of two genes). The two chromsomes 16 have two alpha globin genes each (for a total of four genes). Hemoglobin protein has two alpha subunits and two beta subunits. Each alpha globin gene produces only about half the quantity of protein of a single beta globin gene. This keeps the production of protein subunits equal. Thalassemia occurs when a globin gene fails, and the production of globin protein subunits is thrown out of balance. (http://www-rics.bwh.harvard.edu/sickle/thal_inheritance.html)

2.4.1 Gene deletion

Gene deletion is not the major cause of β-thalassemia. A 3.4-kb deletion, which removes the entire β-globin gene, has been found in Thailand (Sanguansermsri, 1990; Lynch, 1991) and recently, a β-globin gene deletion of 105 bp was detected in β-thalassemia patients in Southern Thailand (Napparatana, 1995). A 619 bp deletion type of β-thalassemia has also been found in some Asian Indians living in Thailand (Winichagoon, 1990). Point mutations and small deletions or insertions in the nucleotide sequence are the main molecular defects responsible for most β-thalassemia. In Thailand, more than 20 different mutations have been detected, the most common being the 4 bp deletion in codons 41/42 (-TTCT), which accounts for 45% of the β-thalassemia alleles. However, different parts of the country have a different incidence of the common β-thalassemia mutations (Table 2). This probably reflects differences in the ethnic population residing within these regions. For example, a G→-C mutation at IVS–1nt5 occurs in about 4.3% of patients from central Thailand but appears to be much more frequent in the southern Thailand (Napparatana, 1995). Some cases of β-thalassemia are due to novel mutations originally reported in Thai patients. These are the ochre mutation C→A at codon 35; -C at codon 41, and +A at codon 95 (Fucharoen, 1989; Fucharoen, 1991 and Winichagoon, 1992). Unstable Hb was detected in β-thalassemia patients: Hb Khon Kaen (8 bp deletion, codons 123-125) (Fucharoen, 1991).

2.4.2 Non-deletion

The non-deletion thalassemia mutation has been further classified into 6 categories. (Honing and Adams III, 1986) such as defective transcription, RNA processing abnormalities, a mutation involving the translation initiation codon, or premature termination of mRNA translation etc. The globin mRNA is in these various disorders accordingly may be non-functional (Kan *et al.*, 1982), unstable (Maquat *et al.*, 1981), or decreased in amount (Treisman *et al.*, 1983). The group of point-mutation globin variants, which characteristically produce the thalassemia phenotype, represents an additional subset of non-deletion thalassemia. These variants, most of which are

extremely unstable. Most of the known example of non-deletion thalassemia mutations consists of single-nucleotide substitutions. A smaller group of these mutations involve deletions or insertions from one to five nucleotide bases, resulting in a shift in the mRNA reading frame. The splice junction mutation is example involving the β-thalassemia. The process of RNA splicing is the excision ligations reactions are required to convert the primary globin gene transcription products to RNA templates whose triplet base codons are colinear with the amino acid sequences of the globin chains. The exon-intron boundary regions of genes from numerous different organisms have shown a considerable degree of sequence homology (Lerner *et al.*, 1980; Breathnach and Chambon, 1981). The "consensus sequence" of these splice junction regions are apparent invariant requirement for normal splicing is the presence of a GU sequence at the 5' end of an intron and an AG at 3' end (Breathnach *et al.*, 1978). The 5' end of this RNA molecule has a remarkable degree of complementarily of splice junction sequences, and it has been proposed that RNA may direct excision-ligation reactions by a base-pairing mechanism. The excision-ligation reactions are strongly dependent on the presence of specific nucleotide sequences at the splice junctions (Breathnach and Chambon, 1981; Mount, 1982) and changes at these sites would therefore be anticipated to impede the formation of normal mature mRNA. A number of globin gene splice junction mutations have been identified as shown in Figure 6, all of which are expressed as thalassemia. Differentials of single-base substitutions have been identified in the 5' splice site region of the β-globin gene IVS-1. Two of these mutations, which involve the invariant GT splice junction dinucleotide (Figure 8 A, B) cause complete inactivation of the splice site, with utilization of a normally inactive nearby 5' splice site (Treisman *et al.*, 1983; Kazazian *et al.*, 1984a). The mutations illustrated in Figure 8 C and D were found to produce normally processed mRNA as well as two abnormal forms of mRNA, reflecting the activation of alternate splice sites (Treisman *et al.*, 1983).

Table 2. Molecular mechanisms of β-thalassemia in Thailand (references in parentheses; ND= not determined) (Fuchareon and Winichagoon, 1997).

		Central (28,29)		South (27)		Northeast (31)		North (30)	
#	Mutations	n	%	n	%	n	%	n	%
1.	-86, C→G	2	0.5	ND	-	ND	-	ND	-
2.	-28, A→G	35	9.3	16	5.7	1	1.7	4	3.5
3.	Cap +1, A→C	0	0	1	0.4	0	0	ND	-
4.	Codons 8/9, +G	ND	-	1	0.4	ND	-	ND	-
5.	Codons 14/15, +G	1	0.3	ND	-	ND	-	ND	-
6.	Codon 15, -T	1	0.3	0	0	ND	-	ND	-
7.	Codon 15, G→A	0	0	1	0.4	0	0	ND	-
8.	Codon 17, A→T	62	16.5	32	11.3	13	21.7	45	39.8
9.	Codon 19, A→G	11	2.9	43	15.2	0	0	ND	-
10.	Codon 26, G→A (Hb E)	2	0.5	ND	-	9	15.0	ND	-
11.	Codon 26, G→T	ND	-	ND	-	1	1.7	ND	-
12.	Codons 27/28, +C	2	0.5	ND	-	ND	-	ND	-
13.	IVS-I-1, G→T	5	1.3	17	6.0	1	1.7	ND	-
14.	IVS-I-1, G→A	0	0	2	0.7	ND	-	ND	-
15.	IVS-I-5, G→C	16	4.3	53	18.8	0	0	2	1.8
16.	Codon 35, C→A	10	2.7	0	0	0	0	ND	-
17.	Codon 41, -C	3	0.8	4	1.4	ND	-	ND	-
18.	Codons 41/42, -TTCT	156	41.6	85	30.1	19	31.6	45	39.8
19.	Codon 43, G→T	3	0.8	0	0	ND	-	ND	-
20.	Codons 71/72, +A	8	2.1	0	0	8	13.3	ND	-
21.	Codon 95, +A	1	0.3	ND	-	ND	-	ND	-
22.	IVS-II-654, C→T	30	8.0	6	2.1	5	8.3	1	0.9
23.	105 bp deletion	ND	-	1	0.4	ND	-	ND	-
24.	619 bp deletion	4	1.1	ND	-	0	0	ND	-
25.	3.4 kb deletion	4	1.1	12	4.3	0	0	1	0.9
26.	Unknown	19	5.1	8	2.8	3	5.0	15	13.3
	TOTAL	375		282		60		113	

	29	30	
	Gly	Arg	
β IVS-1	GGC	AG ¦	gttggtat
A) Base substitution	GGC	AG ¦	attggtat
B) Base substitution	GGC	AG ¦	tttggtat
C) Base substitution	GGC	AG ¦	gttgctat
D) Base substitution	GGC	AG ¦	gttggcat

Figure 9. The mutation of splice junction. Each of these mutations interferes with RNA processing, resulting in a deficiency of normally spliced globin mRNA (Honig and Adams III, 1986).

2.5 Sickle / β-thalassemia

The sickle syndromes are a group of gene disorders with the common feature of having at least one gene that produces hemoglobin S. The sickle/β-thalassemia is a heterozygote of Hb S gene and β°-thalassemia gene. Generally, there are two types of β-thalassemia, there is no β-globin synthesis, while in β^{+}-thalassemia, the defect leads to the decreased of β-globin synthesis. The Hb S gene is responsible for the structure of hemoglobin molecule in the red blood cells. The inheritance of sickle/β-thalassemia was shown in Figure 10.

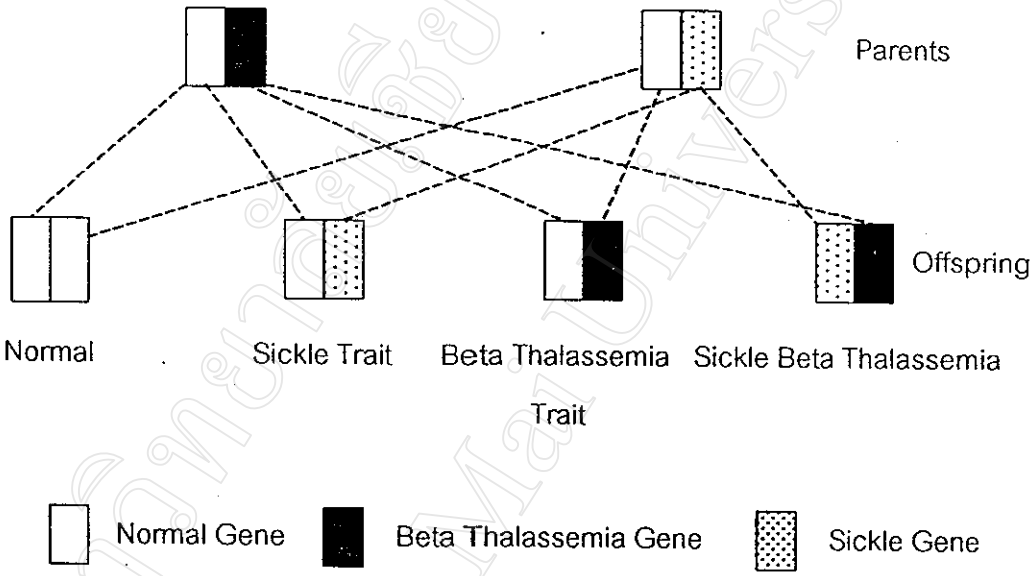

Figure 10. Shown the phenotypic probability of the offspring. As indicated the chance of sickle/β-thalassemia child is one in four.
(http://www-rics.bwh.harvard.edu/sickle/thal_inheritance.html)

2.6 The Polymerase Chain Reaction

The polymerase chain reaction (PCR) is an *in vitro* technique that allows the amplification of a specific deoxyribonucleic acid (DNA) region that lies between two regions of known DNA sequence. Kary B. Mullis invented this technique in 1983. The essential step in each cycle are thermal denaturation of double stranded target molecules, primer annealing to both DNA strands and enzymatic synthesis of DNA (Vosberg, 1989). Primers are short, single stranded DNA molecules between 20 and 30 nucleotides in length, which are complementary to the ends of a defined sequence of DNA template. A DNA polymerase in the presence of deoxynucleoside triphosphates (dNTPs) extends the bond primers on single-stranded denatured DNA template under suitable reaction conditions (Newton and Graham, 1994). This results in the synthesis of new DNA strands complementary to the template strands. These strands exist at this stage as double stranded. DNA strand synthesis can be repeated by heat denaturation of the double-stranded DNA, annealing of primers by cooling the mixture and primer extension by DNA polymerase at a temperature suitable for the enzyme reaction. Each repetition of strand synthesis comprises a cycle of amplification. Each new DNA strand synthesized becomes a template for any further cycle of amplification and so the amplified target DNA sequence is selectively amplified cycle after cycle (Figure 11). This technique is capable of synthesis over a million copies of a specific target DNA sequence in a few hours, significantly facilitating all subsequent analytic procedures. The products of a successful first round of amplification are heterogeneously sized DNA molecules, whose lengths may exceed the distance between the binding sites of the two primers. In the second round, these molecules generate DNAs of defined length that will accumulate in an exponential fashion in later rounds of amplification and will form the dominant products of the reaction. Although longer molecules continue to be produced from the original template DNAs in every round, they accumulate only at a linear rate and therefore do not contribute significantly to the final mass of target sequence (Sambrook *et al.*, 1989). This results in the exponential accumulation of the specific target fragment at approximately 2^n, when n is the number of cycles.

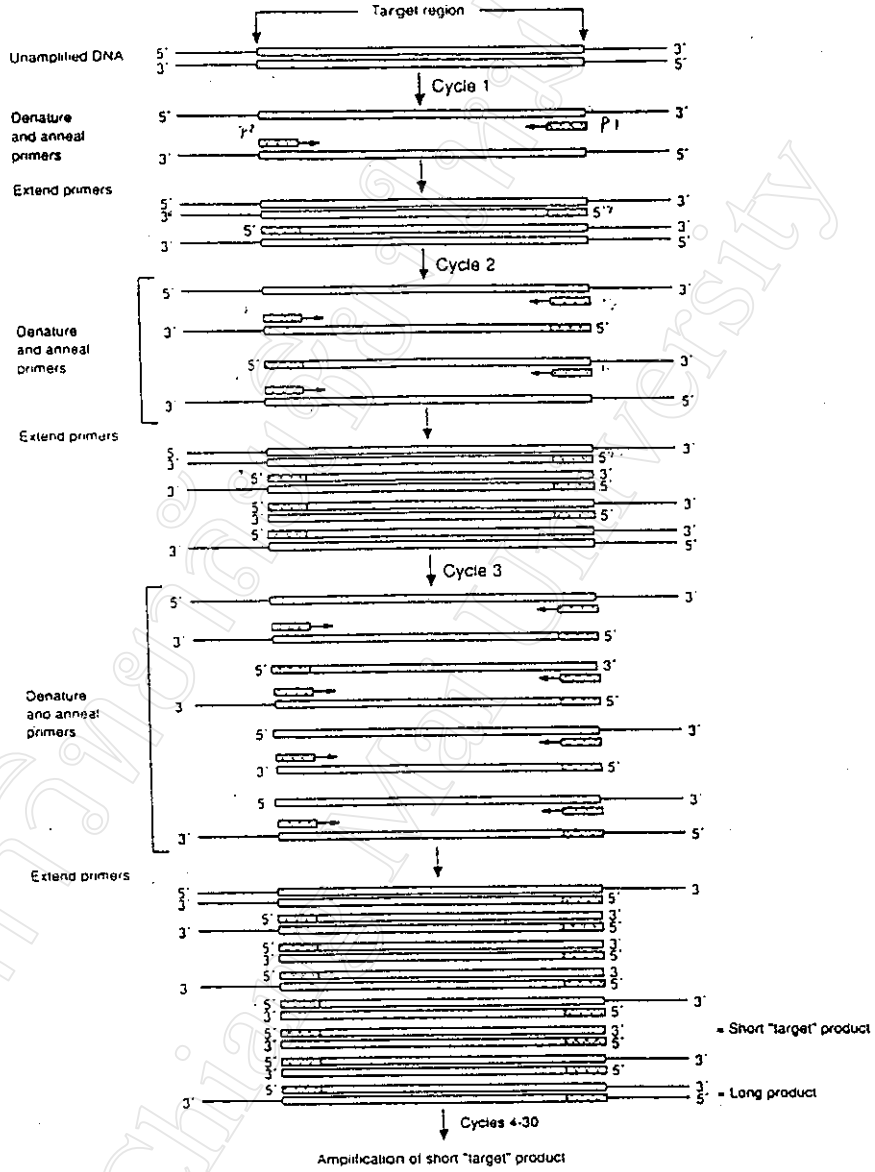

Figure 11. The polymerase chain reaction. PCR is a cycling process; with each cycle the number of DNA targets doubles. The strands in the targeted DNA are separated by thermal denaturation and then cooled to allow primers to anneal specifically the target region. DNA polymerase is then used to extend the primers in the presence of the four dNTPs and suitable buffer. In this way duplicates of the original target region are produced and this 'cycle' is normally repeated for 20-40 cycles. The short 'target' products, which increase exponentially after the fourth cycle, and the long products, which increase linearly, are shown (Newton and Graham, 1994).

2.7 The thermal cycle sequencing

Thermal cycle sequencing, first described in 1989, is an integration of two technologies; the dideoxy chain termination method of Sanger (Sanger *et al.*, 1977) and thermal cycling methodology as used in the PCR technique (Saiki *et al.*, 1985). Cycle sequencing in which a small number of template DNA molecules are respectively utilized to generate a sequencing ladder. A dideoxy sequencing reaction mixture (template, primer, dNTPs, ddNTPs and a thermostable DNA polymerase) is subjected to 20 to 30 cycle of denaturation, annealing and synthesis steps, similar to PCR. In this manner, linear amplification of the sequencing products occurs, that allowing much less template DNA to be used than is usually required. Finally, the rapid thermal cycling helps prevent sequencing problems due to reannealing of linear double-stranded templates such as PCR products. Because a thermostable DNA polymerase is required, this method offers the option of performing the elongation reaction at a high temperature to destabilize secondary structures that may be present in the DNA template. Thermal cycle sequencing is compatible not only with traditional radiolabeling methods, nonradioactive labeling methods such as chemiluminescence (using 5'-end-biotinylated primer) or fluorescence (in automated sequencers using fluorescently end-labeled primers or ddNTPs; Sears *et al.*, 1992). Recently for thermal cycle sequencing reaction various protocols have been developed such as Carothers *et al.*, 1989 and Sears *et al.*, 1992.

2.8 Thermal cycle sequencing with dRhodamine dye terminators and Ampli*Taq*® DNA polymerase, FS enzymes

The high resolution of DNA sequencing has generated a considerable interest in DNA sequencing protocols for the analysis of PCR products (Gibbs, 1990). This cycle sequencing protocols relies on the use of Ampli*Taq*® DNA polymerase, FS. This enzyme is a variant of *Thermus aquaticus* DNA polymerase that contains a point mutation in the active site and also has a second mutation in the amino terminal domain that virtually eliminates the 5'→3' nuclease activity of Ampli*Taq*® DNA polymerase. The enzyme has been formulated with a thermally stable inorganic pyrophosphatase to eliminate problems associated with pyrophosphorolysis.

Advantages of the cycle sequencing protocol over traditional sequencing methods are less hands-on operation, no alkaline denaturation step required for double-stranded DNA, same protocol for both single-and double-stranded templates, less starting template needed and more reproducible results. PE Applied Biosystems designed the dye terminator dichlororhodamine (dRhodamine).

2.9 ABI PRISM 310 Genetic Analyzer

The ABI PRISM 310 genetic analyzer detects DNA fragment that labeled with different dyes, which are used to identify the A, G, C and T extension reaction. Each dye emits light at a different wavelength when excited by laser light; thus, all four colors (and there are all four reactions) can be detected and distinguished in a single gel lane. The ABI PRISM 310 genetic analyzer performs acrylamide separations using capillary electrophoresis technology. The software operates with the ABI PRISM 310 genetic analyzer to determine size and quantify DNA fragments by automated fluorescence detection. This system eliminates the need for radioactivity and post-electrophoresis gel handling. The precise quantitation allowed by fluorescence detection is a valuable feature when measuring gene dosage or studying patterns of inheritance. The resolution is sufficient to accurately analyze molecular lengths up to 5,000 base pairs. But in this study the resolution is sufficient to accurately analyze molecular lengths 300-700 base pairs. The result of an experiment can be displaed as electropherograms, as tubular data, or as a combination of both. Electropherograms show fluorescence as a function of time or size. ABI PRISM 310 electropherograms represent a single injection for the data.

In this study, the three exons of the β-globin gene were amplified by PCR and each amplicon was cycle sequenced. The literature about sequencing of β-globin gene exons:

Chang *et al* (1999): study of DNA sequncing of β-globin gene in Taiwanse family. The results showed a G⟶A substitution at codon 7, Hb Siriraj. This Hb was discovered in Thai family in 1965 by Tuchinda *et al.* Moreover, Hb Siriraj was created by *Mbo*II cutting site.

Fucharoen *et al.* (1991): study of direct DNA sequencing of β-globin gene in a 3-year old patient from northern Thai that found a novel variant, Hb Khon Kaen. The mutation is a 8-bp deletion, codon123-125 in exon 3 of β-globin gene.

Hutt *et al.* (1996): study of DNA sequencing of β-globin gene in a Southeast Asian family showed that the mutation is AAA→ACA in codon 132 of exon 3, corresponding to the amino acid substitution Lys→Thr.

Hutt *et al.* (1996): study of DNA sequencing of β-globin gene in hemoglobin variants showed that two amino acid substitution affecting one globin chain. The mutations are GAG→AAG in codon 26 of exon 1, corresponding to the amino acid Glu→Lys, the known substitution of Hb E and GAA→CAA in codon 121 of exon 3, consistent with the known Glu→Gln substitution of Hb D-Punjab.

Okada *et al.* (1995): analyzed the hemoglobin of a girl with β–thalassemia and those of her immediate family. DNA sequencing of the cloned β-globin gene found a point mutation at the IVS-1 position 1(G→T). This is the first reported Japanese case.

Itchayanan *et al* (1999): study of direct DNA sequencing of β-globin gene showed that the mutation is GAA→GCA in codon 22 of exon 1, corresponding to the amino acid substitution Glu→Ala substitution of Hb G-Coushatta.